



DECEMBER 2024

BIPARTISAN HOUSE TASK FORCE REPORT ON ARTIFICIAL INTELLIGENCE

Guiding principles, forward-looking recommendations,
and policy proposals to ensure America continues to
lead the world in responsible AI innovation

118TH CONGRESS



Table of Contents

Table of Contents	i
Letter to Speaker Johnson and Democratic Leader Jeffries	iii
About the Bipartisan House AI Task Force in the 118th Congress	iv
Leading AI Progress: Policy Insights and a U.S. Vision for AI Adoption, Responsible Innovation, and Governance	v
Philosophy and Principles	vi
Overview of Findings and Recommendations	ix
Government Use	1
Key Findings in Government Use	17
Recommendations in Government Use	19
Federal Preemption of State Law	23
Key Findings in Federal Preemption of State Law	29
Recommendations in Federal Preemption of State Law	30
Data Privacy	31
Key Findings in Data Privacy	37
Recommendations in Data Privacy	38
National Security	39
Key Findings in National Security	52
Recommendations in National Security	53
Research, Development, and Standards	54
Key Findings in Research, Development, and Standards	71
Recommendations in Research, Development, and Standards	73
Civil Rights and Civil Liberties	78
Key Findings in Civil Rights and Civil Liberties	84
Recommendations in Civil Rights and Civil Liberties	85
Education and Workforce	87
Key Findings in Education and Workforce	106
Recommendations in Education and Workforce	107
Intellectual Property	111
Key Findings in Intellectual Property	135
Recommendations in Intellectual Property	136

Content Authenticity	137
Key Findings in Content Authenticity	152
Recommendations in Content Authenticity	153
Open and Closed Systems	155
Key Findings in Open and Closed Systems	160
Recommendations in Open and Closed Systems	161
Energy Usage and Data Centers	162
Key Findings in Energy Usage and Data Centers	173
Recommendations in Energy Usage and Data Centers	174
Small Business	176
Key Findings in Small Business	182
Recommendations in Small Business	183
Agriculture	185
Key Findings in Agriculture	196
Recommendations in Agriculture	198
Healthcare	200
Key Findings in Healthcare	218
Recommendations in Healthcare	219
Financial Services	221
Key Findings in Financial Services	236
Recommendations in Financial Services	237
Appendix I: AI Task Force Members	239
Appendix II: AI Task Force Events	242
Appendix III: Key Government Policies	246
Appendix IV: Areas for Future Exploration	248
Appendix V: Overview of AI Technology	249
Appendix VI: Definitional Challenges of AI	251
Acknowledgments	253

Bipartisan House Task Force on Artificial Intelligence

The Honorable Mike Johnson
Speaker
United States House of Representatives
Washington, DC 20515

The Honorable Hakeem Jeffries
Democratic Leader
United States House of Representatives
Washington, DC 20515

Dear Speaker Johnson and Leader Jeffries:

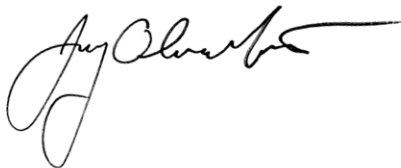
We, the Co-Chairs of the Bipartisan Artificial Intelligence Task Force, submit to you our key findings in this report.

Although artificial intelligence (AI) is not a new concept, breathtaking technological advancements in the last few years have made AI the focus of numerous policy discussions. AI has tremendous potential to transform society and our economy for the better and address complex national challenges. From optimizing manufacturing to developing cures for grave illnesses, AI can greatly boost productivity, enabling us to achieve our objectives more quickly and cost-effectively. Nevertheless, we also recognize that AI can be misused and lead to various types of harm.

This report highlights America's leadership in its approach to responsible AI innovation while considering guardrails that may be appropriate to safeguard the nation against current and emerging threats. You charged twenty-four members, twelve Republicans and twelve Democrats, with developing a U.S. vision for AI adoption, innovation, and governance. The AI Task Force gathered information on salient AI issues from domain experts in industry, government, civil society, and academia to provide 66 key findings 85 recommendations. In summary, this report encapsulates a targeted approach that balances the need to promote vibrant AI innovation while safeguarding Americans from potential harms as we enter an era of widespread adoption of AI.

We thank you for establishing the AI Task Force and are eager for this report to inform future congressional policymaking.

Sincerely,



Jay Obernolte
CHAIRMAN



Ted W. Lieu
CO-CHAIRMAN

About the Bipartisan House AI Task Force in the 118th Congress

The bipartisan AI Task Force was created by Speaker Johnson and Democratic Leader Jeffries on February 20, 2024. The AI Task Force is led by co-chairs Jay Obernolte (R-CA) and Ted Lieu (D-CA) and comprises twenty-four members, twelve Republicans and twelve Democrats. The AI Task Force members are drawn from twenty committees to ensure comprehensive jurisdictional responsibilities over the numerous AI issues that we addressed and to benefit from a range of different insights and perspectives.

A full list of Task Force members and the committees they represent is included in [Appendix I](#).

Throughout 2024, the AI Task Force convened to investigate dozens of issues at the heart of how AI intersects with numerous policy areas. The AI Task Force held multiple hearings and numerous roundtables and engaged with over one hundred experts, including business leaders, government officials, technical experts, academics, legal scholars, and other domain specialists. These experts generously offered their insights, suggestions, and comments spanning a range of viewpoints.

This approach allowed each issue to be comprehensively explored from various perspectives. A multifaceted approach to policy analysis will better prepare the decision-makers who address the complex AI challenges that confront our nation and will continue to affect public policy.

A full list of experts and a list of the events the Task Force convened is included in [Appendix II](#).



DATA PRIVACY

Background

As AI systems amass and analyze vast amounts of data, there are increasing risks of private information being accessed without authorization. Training algorithms identify patterns within the data and produce a set of instructions or a model that can be used with new data.¹ AI models are often trained on diverse datasets that include text from books, websites, and other digital sources, some of which may contain personal or sensitive information. AI systems can also be deployed in sensitive contexts, including healthcare settings, that rely on sensitive data. When users interact with some AI systems, especially generative AI systems, they can inadvertently reveal private or confidential information stored and processed by the AI. Each of these situations has provoked significant concerns regarding the data privacy challenges associated with AI.

Thoughtful and effective data privacy policies and protections will support consumer confidence in the responsible development and deployment of AI systems. While the House AI Taskforce has endeavored to examine data privacy in the context of AI, further exploration of this issue is warranted. Committees with jurisdiction over data privacy should continue to invest time and resources in examining these problems and proposing solutions for the American people.

Advanced AI Systems Require Increasing Amounts of Data

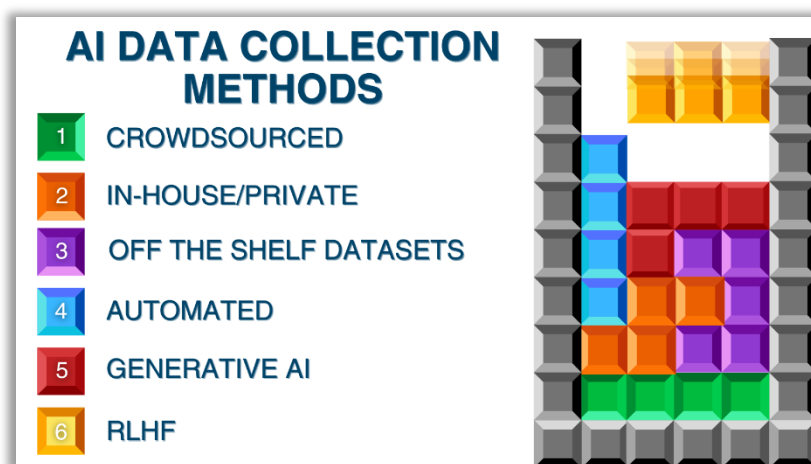
If the data used for training is too small or of poor quality, the model may perform suboptimally. Using large quantities of data from multiple diverse sources generally allows the trained models to perform better.

¹Tursman, Eleanor, et al. "AI 101 - Aspen Digital." Aspen Digital, 20 June 2023, www.aspendigital.org/report/ai-101/#section2.

According to some research, the performance of algorithms benefits significantly from larger training datasets.^{2,3} In the decades since, the amount of digital data that could be used for training algorithms has increased dramatically.⁴ To be sure, a growing volume of research suggests smaller datasets may also play a role in improving AI performance.⁵

Data available to train AI models is collected and licensed in various ways. Some companies use a combination of internal and external data.⁶ Other firms, such as those deploying large language models (LLMs) and foundation models, mainly rely on data acquired (“scraped”) from the internet.

Web scraping is a process by which data is copied from the internet. Some companies package, process, and label the scraped data for sale, while others release open-source data sets.⁷ There is a voluntary standard used by many websites to indicate that they should not be scraped, and other companies have added such a stipulation to their terms of service. Unfortunately, these clearly stated requests are often ignored⁸ and there are a growing number of disputes and litigation over scraping issues involving AI companies.



Source: AI Multiple Research - Top 6 Data Collection Methods for AI & Machine Learning

² Banko, Michele, and Eric Brill. “Scaling to Very Very Large Corpora for Natural Language Disambiguation.” ACL Anthology, 2001, aclanthology.org/P01-1005.pdf.

³ Kaplan, Jared, et al. “Scaling Laws for Neural Language Models.” arXiv, 23 Jan. 2020, <https://arxiv.org/abs/2001.08361>.

⁴ Roded, Tal, and Peter Slattery. “What Drives Progress in AI? Trends in Data.” FutureTech, 19 March 2024, futuretech.mit.edu/news/what-drives-progress-in-ai-trends-in-data.

⁵ Li, Kangming, et al. “Exploiting redundancy in large materials datasets for efficient machine learning with less data.” Nature Communications, 10 Nov. 2023, <https://www.nature.com/articles/s41467-023-42992-y>.

⁶ Brown, Sara. “Why External Data Should Be Part of Your Data Strategy.” MIT Sloan School, 18 Feb. 2021, mitsloan.mit.edu/ideas-made-to-matter/why-external-data-should-be-part-your-data-strategy.

⁷ Newman, Marissa, and Aggi Cantrill. “A High School Teacher’s Free Image Database Powers Ai Unicorns.” Bloomberg, 24 Apr. 2023, www.bloomberg.com/news/features/2023-04-24/a-high-school-teacher-s-free-image-database-powers-ai-unicorns.

⁸ Paul, Katie. “Exclusive: Multiple AI companies bypassing web standard to scrape publisher sites, licensing firm says.” Reuters, 21 June 2024, <https://www.reuters.com/technology/artificial-intelligence/multiple-ai-companies-bypassing-web-standard-scrape-publisher-sites-licensing-2024-06-21/>

Companies are also turning to their own users' data to train AI systems. Google allegedly scraped Google Docs and Gmail for data to train AI tools.⁹ Users may also transmit their personal or company data via queries provided to AI models that are hosted or otherwise controlled by a third party, like an AI company. Meta and X have changed their privacy policies to allow for training AI models on the platforms' data.^{10,11}

More companies are updating their privacy policies in order to permit the use of user data to train AI models.¹² Meta faces legal challenges in eleven European countries over its plans to use users' personal data to train AI models.¹³

The Federal Trade Commission (FTC) has addressed the matter, stating that it may be unfair or deceptive for a company to adopt more permissive data practices but only inform consumers of this change through an amendment to its privacy policy.¹⁴

In response to these concerns, some companies are turning to privacy-enhancing technologies, which seek to protect the privacy and confidentiality of data when sharing it. For example, Apple has used a privacy-preserving technology called differential privacy to analyze Apple users without sharing individuals' information.¹⁵

Similarly, the AI company Anthropic recently partnered with the UK Safety Institute and the PET company OpenMined to test how to utilize secure computation to allow multiple parties to access advanced models and nonpublic data.¹⁶

The growth in data widely available to AI companies may be reaching a plateau.¹⁷ It is unclear how AI developers and researchers will satisfy the need for additional training data. Some are exploring synthetic data, which is created artificially through computer simulations or algorithms.

⁹ Morrison, Sara. "The Tricky Truth about How Generative AI Uses Your Data." Vox, 27 July 2023, www.vox.com/technology/2023/7/27/23808499/ai-openai-google-meta-data-privacy-nope.

¹⁰ Mearian, Lucas. "Meta's Privacy Policy Lets It Use Your Posts to Train Its AI." Computerworld, 21 June 2024, www.computerworld.com/article/2264949/metas-privacy-policy-lets-it-use-your-posts-to-train-its-ai.html.

¹¹ Perez, Sarah. "Elon Musk's X Is Changing Its Privacy Policy to Allow Third Parties to Train AI on Your Posts." TechCrunch, 17 Oct. 2024, techcrunch.com/2024/10/17/elon-musks-x-is-changing-its-privacy-policy-to-allow-third-parties-to-train-ai-on-your-posts/.

¹² Hays, Kali. "A Long List of Tech Companies Are Rushing to Give Themselves the Right to Use People's Data to Train AI." Business Insider, 13 Sept. 2023, www.businessinsider.com/tech-updated-terms-to-use-customer-data-to-train-ai-2023-9.

¹³ Woollacott, Emma. "Meta Faces Legal Complaints Over New AI Training Data Plans." Forbes, 10 June 2024, www.forbes.com/sites/emmawoollacott/2024/06/10/meta-faces-legal-complaints-over-new-ai-training-data-plans/.

¹⁴ Staff in the Office of Technology and The Division of Privacy and Identity Protection. "AI (and Other) Companies: Quietly Changing Your Terms of Service Could Be Unfair or Deceptive." Federal Trade Commission, 13 Feb. 2024, www.ftc.gov/policy/advocacy-research/tech-at-ftc/2024/02/ai-other-companies-quietly-changing-your-terms-service-could-be-unfair-or-deceptive.

¹⁵ "Differential Privacy Overview." Apple, 2 Nov. 2017, images.apple.com/privacy/docs/Differential_Privacy_Overview.pdf.

¹⁶ "Interviewing Andrew Trask on How Language Models Should Store (and Access) Information." Interconnects, 10 Oct. 2024. www.interconnects.ai/p/interviewing-andrew-trask.

¹⁷ Xu, Tammy. "We Could Run out of Data to Train AI Language Programs ." MIT Technology Review, 24 Nov. 2022, www.technologyreview.com/2022/11/24/1063684/we-could-run-out-of-data-to-train-ai-language-programs/.

Synthetic data can be used as an alternative or supplement to real-world data, particularly when real-world data of the appropriate form is unavailable or has already been exhausted.¹⁸ Synthetic data allows for exploring new possibilities because such data can be designed to represent hypothetical situations beyond what existing real-world data represents.¹⁹ Synthetic data also offers privacy-enhancing benefits, given that it does not include information on real individuals. Unfortunately, since it does not truly represent actual measurements, synthetic data may lack the complexity and nuances of real-world data. Accordingly, models trained on synthetic data may be unable to perform well in various real-world scenarios, and an overreliance on synthetic data may lead to technical complications like model collapse.²⁰

Privacy Harms From AI

Americans are vulnerable to several privacy harms. The full breadth of privacy harms is difficult to estimate because they are so varied and can encompass different but related concerns. Nevertheless, to clarify the policy issues, the following types of privacy harms are frequently referenced by stakeholders:

- *Physical harms* result in bodily injury or death. For example, a man purchased personal data about Amy Boyer from New Hampshire, including the address of Boyer's employer. The man fatally shot her where she worked.
- *Economic harms* involve monetary losses or other losses of value. Identity thieves steal personal data and use it to conduct fraudulent transactions in victims' names, including opening credit card accounts and accruing debt that damages the victims' credit history.

Increasingly, these thieves target public school districts and steal the identities of children. The credit records of minors can be exploited for years before the victims even discover it. Celeste Gravatt is one of the thousands of parents who had her children's data stolen as part of a Minneapolis Public Schools cyberattack. She locked their credit accounts but remains worried.

- *Emotional harms* result from emotional distress from information being released about someone without their knowledge or consent. These harms form the basis of many privacy torts, such as intrusion upon seclusion, trespass, and more.

¹⁸ IBM. "What Is Synthetic Data?" IBM, 2024, www.ibm.com/topics/synthetic-data.

¹⁹ Bozzella, Kim. "The Pros And Cons Of Using Synthetic Data For Training AI." Forbes, 20 Nov. 2023, www.forbes.com/councils/forbestechcouncil/2023/11/20/the-pros-and-cons-of-using-synthetic-data-for-training-ai/.

²⁰ Shumailov, Ilia, et al. "AI models collapse when trained on recursively generated data." Nature, 24 July 2024, <https://www.nature.com/articles/s41586-024-07566-y>.

- *Reputational harms* involve injuries to an individual's reputation and standing in the community, such as lost business, employment, or social status. For example, Murray Dowey was the target of "sextortion," online blackmail based on the threat of exposing his intimate images. Dowey tragically took his own life earlier this year.²¹
- *Discrimination harms* involve disadvantaging people based on characteristics like sex, race, age, religion, or political affiliation. They can thwart people's ability to obtain jobs, secure insurance, and find housing.
- *Autonomy harms* involve subverting or impairing an individual's autonomy. For example, some bad actors use "dark patterns" or design features used to deceive or manipulate users.

There are many examples of AI systems exacerbating privacy harms. Synthetic content can duplicate someone's likeness without their consent. Facial recognition systems can enable pervasive tracking of people in public places. Advanced AI systems, such as LLMs, have been found to inadvertently leak personally identifiable information if not properly configured or protected.²² Further, AI systems have been shown to infer sensitive information about someone,²³ even from legally obtained and deidentified data,²⁴ in some cases inadvertently revealing personal attributes such as political views or sexual orientation. In one case, a major retailer's system predicted a shopper was pregnant and accidentally revealed that information to her father.²⁵

American's Privacy Protections Vary

Currently, there is no comprehensive U.S. federal data privacy and security law. However, there are several federal privacy laws focused on various sectors or use cases, such as child privacy or health information. States have also acted. To date, nineteen U.S. states have enacted their own state privacy laws with varying standards.²⁶

²¹ Chigozie Ohaka, et al., "Stop terrorizing children with sextortion, say parents." *BBC*, November 2024. <https://www.bbc.com/news/articles/cz6jywx37dlo>

²² Yan, Biwei, et al. "On Protecting the Data Privacy of Large Language Models (LLMs): A Survey." arXiv, 8 March 2024, arxiv.org/abs/2403.05156.

²³ Crețu, Ana-Maria, et al. "Interaction data are identifiable even across long periods of time." *Nature Communications*, 25 Jan. 2022, <https://www.nature.com/articles/s41467-021-27714-6>

²⁴ Na, Lingyuan, et al. "Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets From Which Protected Health Information Has Been Removed With Use of Machine Learning." *JAMA Network Open*, 21 Dec. 2018, <https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2719130>

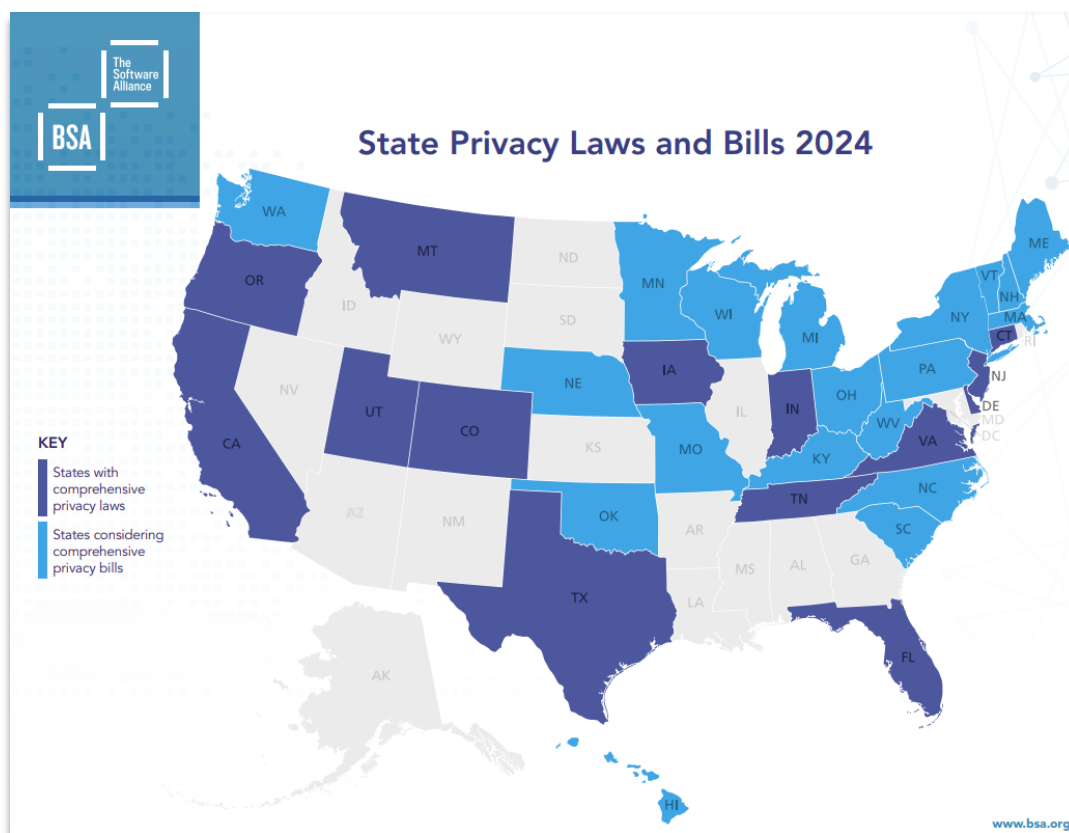
²⁵ Hill, Kashmir. "How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did." *Forbes*, 11 Aug. 2022, www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/.

²⁶ Kibby, C. "U.S. State Privacy Legislation Tracker." Resource Center, iapp, 4 Nov. 2024, iapp.org/resources/article/us-state-privacy-legislation-tracker/. Some states, such as New York and Colorado, have enacted AI-related legislation as well, while others like California are seeking to use their state data privacy and security laws to regulate AI.

One salient example of state action is data breach notification laws; in the absence of a federal standard, each state moved forward to create its own.²⁷ State laws have created a patchwork of rules and regulations with many drawbacks. Consumers can be confused about the extent of privacy protections, gaps exist in data privacy protections, and businesses can face increased compliance burdens and uncertainty.

Companies that fail to comply with each existing privacy law risk a myriad of lawsuits from state regulators and individuals. As such, private companies conducting business in multiple states must track and monitor changes to state laws, which is often a challenging task for smaller businesses with fewer resources. If businesses are forced to comply with multiple state laws, it can inhibit business expansion, hiring, and the ability to develop and deploy new technologies.

Federal legislation that preempts state data privacy laws has advantages and disadvantages.²⁸ The complexities of federal preemption are discussed further in the chapter on [Federal Preemption of State Law](#).



Source: BSA | The Software Alliance - State Privacy Bills and Laws Map

²⁷ "Data Breach Notification Laws by State." IT Governance, www.itgovernanceusa.com/data-breach-notification-laws.

²⁸ Mulligan, Stephen P., and Chris D. Linebaugh. "Data Protection and Privacy Law: An Introduction." Congressional Research Service, 12 Oct. 2022, crsreports.congress.gov/product/pdf/IF/IF11207.

Key Findings

AI has the potential to exacerbate privacy harms.

AI is inherently linked to issues of data: how to obtain large amounts of data, how to analyze data for patterns, and how to use those patterns to make predictions. Developers and users of AI can intentionally or unintentionally cause or exacerbate data privacy harms related to each of these facets.

Americans have limited recourse for many privacy harms.

Many businesses are generally unrestricted in the types of sensitive information they can collect from Americans, how they can use that information, who they can transfer or sell it to, and how long they can retain it. While state laws have started to address these concerns, many Americans have limited rights or recourse when faced with encroachments on their privacy.

Federal privacy laws could potentially augment state laws.

Federalism has been a controversial issue for federal data privacy laws because of its complexity. Congress could adopt a comprehensive system for data protection by expressly preempting state laws related to data privacy. Alternatively, Congress could preserve state laws in some ways but preempt them in others. Another option is for Congress to pass a law that preempts state legislation but still enables states to enforce the federal standard. Congress also has the option to leave state schemes intact in conjunction with a federal scheme.

Recommendations

Recommendation: Explore mechanisms to promote access to data in privacy-enhanced ways.

Access to privacy-enhanced data will continue to be critical for AI development. The government can play a key role in facilitating access to representative data sets in privacy-enhanced ways, whether through facilitating the development of public datasets or the research, development, and demonstration of privacy-enhancing technologies or synthetic data. Congress can also support partnerships to improve the design of AI systems that consider privacy-by-design and utilize new privacy-enhancing technologies and techniques.

Recommendation: Ensure privacy laws are generally applicable and technology-neutral.

Congress should ensure that privacy laws in the United States are technology-neutral and can address many of the most salient privacy concerns with respect to the training and use of advanced AI systems. Congress should also ensure that general protections are flexible to meet changing concerns and technology and do not inadvertently stymie AI development.